DOCUMENT RESUME

ED 142 217 IR 005 076

TITLE	Elements of Information and Information Processing for Teachers in Secondary Schools.
INSTITUTION	International Federation for Information Processing, Geneva (Switzerland). Working Group on Secondary School Education.
PUB DATE	76
NOTE	39p.; For related documents, see ED 077 209 and ED 058 889
EDRS PRICE	MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS	Computer Oriented Programs; *Conceptual Schemes; *Information Processing; Information Systems; *Information Theory; Information Utilization; Models; *Problem Solving; Secondary School Teachers; Teacher Education

ABSTRACT

The efficient use of information depends on the way it is organized. Haphazard information may serve some useful purpose, but for a particular purpose, information needs to be processed into another structure which more easily fulfills that purpose. This booklet examines the concepts of information and information processing with the intention of leading on to a study of the way in which problems can be formulated so that a computer can assist in the solving process. It discusses: (1) what information is, (2) how it is represented and coded, (3) how it can be organized, and (4) how it can be processed. Appendices describe additional concepts and suggest topics suitable for group discussion. (Author/DAG)

ELEMENTS OF INFORMATION AND INFORMATION PROCESSING FOR TEACHERS IN SECONDARY SCHOOLS

This document has been prepared by the IFIP Working Group on Secondary School Education (WG 3.1), a Group set up by the IFIP Technical Committee on Education (TC 3).

The following preople in particular contributed to this booklet and their addresses are given on the final page:

Wm. F. Atchison	U.S.A.
A. Berger	Austria
G.C. Bonham	Canadá
U. Bröndum	Denmark
R.A. Buckingham	United Kingdom
E. Ga. Camarero	Spain
Sylvia Charp	U.S.A.
J. Hebenstreit	France
D.C. Johnson	U.S.A.
R.E.J. Lewis	United Kingdom
J.F. Louw	South Africa
F.B. Lovis	United Kingdom
T. Malmberg	Sweden
E.D. Tagg	United Kingdom
J.D. Tinsley	United Kingdom
D.H. Wolbers	Netherlands

The Working Group is also indebted to many other individuals for their helpful suggestions. Special acknowledgement should be made to the CITEMA Foundation through whose generosity it was possible to arrange a meeting of the Group in Madrid in February 1975 and to print this booklet.

Previous booklets by WG 3.1

Computer Education for Teachers in Secondary Schools:

1. An Outline Guide (September 1971)
2. Aims and Objectives in Teacher Training (October 1972)

CONTENTS

PREFACE

This is the third of a series of booklets which the IFIP Working Group on Secondary Education is developing to help those responsible for planning courses for the training of teachers. The first booklet was an 'Outline Guide', essentially a summary of those topics in which computers, or the basic ideas which computers have brought to the fore, are important to teachers of all subjects. As the Outline Guide could not contain sufficient explanation of the fundamental concepts to be of practical value to the less experienced lecturer, further booklets of a more detailed nature are being prepared on more specific themes.

The second in the series described how computers can play a part in the achievement of educational goals. This one explores the concepts of information and information processing with the intention of leading on to a study of the way in which problems can be formulated so that a computer can assist in the solving process. These concepts are among the most difficult likely to be met with in this series, ard are of vital importance in the context of human beings using computers.

Suggestions on the content of this booklet will be welcomed, especially from people who have had practical experience in methods of presenting these topics. These contributions will be incorporated in further revisions of this guide and should be sent to

Wm. F. Atchison
Dept. of Computer Science
University of Maryland
College Park,
Maryland 20742
U.S.A.

I INTRODUCTION

This booklet, which examines the concepts of information and information processing, contains a minimum of direct references to the computer. Its primary aim is to bring to light some of the factors which govern the efficient handling of information, which is becoming increasingly important as an element in the quality of human life.

The efficient use of information depends on the way it is organised. To illustrate this let us consider one of the problems which arises when a manually-operated telephone service is installed in a large city. The simplest way to compile a directory of subscribers' names, addresses and numbers would be to prepare a sequential list as and when the subscribers enroll. Such a directory would soon prove cumbersome, for it could well take a person some fifteen seconds to find one entry among 60. As soon as, say, 600 subscribers have joined, then 10 operators would be needed to provide a satisfactory service. If the number of subscribers should rise to 6 million, such an elementary solution would clearly be impracticable.

This is a problem in information retrieval. Although any required number should be present in the list, the difficulty lies solely in finding it. The solution to the problem lies in changing the existing structure of the information into another structure which suits the purpose of the directory better. This is usually done by a form of processing which alters the list into an alphabetical order of names. The new directory now serves one purpose well — if the enquirer knows the name of the person he wishes to call he will have little trouble. But suppose he knows only the number and wishes to find the name and address of a subscriber. The alphabetical list may then be even less efficient than the original list and the information will need to be subjected to different processing in order to obtain a third list in strict numerical order.

This simple example illustrates the basic principles which lie behind information structure and processing. Haphazard information may well serve some useful purposes, but because its shape or structure is random, or at least unrelated to the particular application, its usefulness will certainly be very restricted. It becomes either laborious to use (as for 600 subscribers) or quite impracticable (for 6 million subscribers). For a particular purpose, information needs to be processed into another structure which more easily fulfills that purpose.

Although computers are being increasingly used to change one structured form of information into other more useful structures, it is important to realise that in doing this the computer is not fulfilling a new function. It is merely one of many information processing devices, including, of course, man himself. Looked at in this way, it seems inadvisable to study the computer in insolation, without considering the common purpose which it shares with all other information processing devices.

When we look closely at information structure and information processing as basic concepts is becomes clear that they are fundamental to everyday life and have long been used without any thought being given to the relevance of computers. It follows that these are concepts which are not tied to computer-based applications and a preliminary study of them should not be confined to such applications.

Indeed at our present state of mental development we spend much time in processing information without pausing to reflect on what we are doing. Only when the processing proves to be unsuccessful are we likely to be disturbed and forced to consider what has gone amiss. This may lead us into a closer analysis of the stages of processing and thence to discover a solution to the problem. On such occasions it will be surprising to find just how much has been taken for granted.

If we decide to use a computer in solving a problem then we must give conscious attention to the data which is available and relevant, to the form of the result which is required, and to the specification of the processing which will have to be carried out. These stages have not been changed fundamentally by selecting a computer as the tool, for they are essentially the same whatever processing device is used.

Returning now to the contention that we should first study information structure and processing without specific reference to the computer, what relevance would such a study have to the school classroom? The answer is that it is wholly relevant, for the simple reason that all successful teaching must involve some form of information processing. If the pupil's store of information is not changed in some way by the teaching he receives then he does not learn. The immense variety of topics which the teacher may call upon to illustrate how we process information is indicated in Appendix II.

II. INFORMATION

The word *information* has different meanings in everyday life depending on the context in which the word is used. In our present context of information processing, information is understood to be the *meaning* which man attaches to *data*. Terms like data, symbol and character string are often used synonymously*. Items or collections of data are not in themselves information but only representations of it, capable of being communicated or manipulated. The term information processing can be used to describe a sequence of activities involving both man and machine:

1. A body of information is first represented as a set of data having an appropriate structure, using an agreed method of representation.
2. When necessary the data items are coded according to a fixed set of man-made rules, and then processed by means of an automatic device, or by man himself, or by a combination of man and machine. Subsequently decoding may be carried out. Throughout these stages it is not necessary to attach any meaning to the data. (The processing may include automatic collection of data by a machine, operations on that data and consequent automatic control of machines).

---

*We should refer to specific definitions of the terms used above as given in the 'IFIP Guide to concepts and terms in data processing', ed I.H. Gould, North-Holland, Amsterdam, 1971. They are as follows:

DATA: A representation of facts or ideas in a formalised manner capable of being communicated or manipulated by some process.

INFORMATION: In *automatic data processing* the meaning that a human expresses by or extracts from *data* by means of the known conventions or representation used.

DIGITAL REPRESENTATION: Representation by means of *characters*, that is, the designation of one out of a finite number of alternatives by a character or a group of characters.

CODE: A mapping of the *characters* of one *character set* (known as the SOURCE SET) on to those of another (known as the TARGET SET), according to a set of rules known as the CODING RULES. (For example, the Morse code 'maps' each letter or number on to a particular combination of dots and dashes.)

Other important concepts are considered in Appendix I.

3. Finally man interprets the processed data, giving it meaning.

These three activities therefore transform given information into new information. The choice of representation, and of coding rules, depends largely on the nature of the processor used and also to some extent on the ease with which meaning can be attached to the data before and after the processing stage.

Only the second activity is known as data processing, which is a term used to describe the manipulation (changing, sorting, retrieving, etc.) of character strings according to a fixed set of procedural rules known as an algorithm. For example, a string of numbers defined to be a set of grades may be averaged and a new number — the average grade — obtained. The sequence of operations (the processing) executed to give the average does not depend on the meaning of the different numbers involved; thus the sequence would be the same if the string of numbers is defined to be children's ages. What would change is the interpretation of the result which would then be the average age instead of the average grade.

Other forms of processing may involve the substitution of one number for another if an amendment is required, or the addition of further numbers as more results come to hand.

SYMBOLS AND THEIR MEANING

Man has always been concerned with information and with its transmission. One of the basic ways in which information can be communicated or transmitted from one person to another is by means of non-verbal signals. For example, a vote on a certain question in a classroom — perhaps the selection of a class project — may be indicated by a show of hands. But in a different context the same show of hands may indicate that several students have a question they wish to ask. Thus one signal can have different meanings in different contexts and it is the teacher who must interpret the intended meaning by the context in which it occurs.

However, of the context cannot be precisely determined then an agreed set of rules for interpreting the signal must be established between the sender and the receiver. If a boat is in shallow waters and there is a man in the front looking out for dangerous rocks, he must agree upon the meaning of this signals with the man who is steering; it must be clear whether his hand signals

mean that he is pointing to a rock or pointing in the direction in which the boat must be steered. From this examples it can be seen that a signal presented in context has meaning and hence is information, but, if the signal is not in context or if the rules for its interpretation are not clear, there is no reliable communication.

It can be seen from these examples that the *meaning* of a signal, i.e. the information, depends not on the way the information is transmitted, but on the rules that are established for interpretation. This can be illustrated further by pointing out that the same information could be transmitted by means other than non-verbal signals. This is possible by the use of recorded information — string of letters, numbers and other symbols, previously referred to as character strings — to convey meanings. Thus a string of letters and numbers on a student record card may represent a list of test scores, medical information or a record of attendance. For these character strings to have meaning and so convey information, the context, the method of representation and the coding rules must all be precisely defined.

When processing information we are concerned with the *meaning* of the string of characters before processing and after processing — that is, the *information* before and after processing.

WHAT ARE THE SOURCES OF INFORMATION?

Much information is gained directly or indirectly from personal experience. Someone deciding upon a vacation may make this decision as a result of his own experience of visits to the sea-shore and to other places. Alternatively, the decision may be based on indirect information from a travel film, or from discussion with friends, or as the result of searching through various travel folders. Other information may arise through attention being drawn to an unusual incident, such as an airline strike or some other breakdown in travel facilities.

Systematic observation of one person by another is a frequent means of acquiring information. A trained observer using routine tests and procedures may collect and record data as when a physician examines a patient. Such data is recorded, using an appropriate code, in the physician's file, and provides him with information about his patient's current state of health.

Many situations arise in which data is gathered by automatic means. A seismograph records signals which are interpreted by an observer to give information about earthquakes. Similarly, a ship's radar screen indicates the location of nearby objects. Sometimes data gathered automatically may be used to activate a process, as when a photo-electric cell detects that a beam of light has been interrupted and this causes a door to be opened automatically. A thermostat may switch a heating source off or on as the temperature rises above or falls below certain levels. Most of the adjustments in the automatic piloting of space vehicles are made on the basis of continual processing of data which results in new data which in turn is fed back into the system to produce further action and so on. Such data may of course become informative if the process is being observed and the observer attaches meaning to what he sees. These examples may be thought of as one mechanism communicating with another to produce an appropriate action. Man has designed many systems in which data is collected automatically, is processed and then initiates some activity. This activity in turn may result in new data which if fed back into the system produces further actions. This process of "feedback" is extremely important in information processing systems.

HOW GOOD IS INFORMATION?

In some cases the nature of the source of information makes it difficult for reliable information to be obtained. When a crime has been committed witnesses often disagree about what they saw. There are some situations where the accuracy of the information is beyond the ability of a human being to ascertain directly without special equipment, for example an interval of time in microseconds or the speed of a passing car involved in an accident.

In other cases incorrect information may be obtained through human or mechanical errors at the source. Thus when a fire occurs it is obviously important that the correct address be given to the fire station. Although this may be done automatically by a fire alarm system or may be reported by telephone, errors may arise because of a fault in the automatic system or an incorrect telephone message.

People sometimes tend to think that when information has gone through a highly organised process, such as being printed in a book or newspaper, added authority is given to the information. This authority is probably justified for a standard work of reference, such as a dictionary where many experts have

checked each entry. However, it remains true that if the source of information is inaccurate or unreliable then no amount of ingenuity will produce correct information. Nevertheless procedures exist which can be used to detect certain types of error and sometimes to make corrections.

There is not always a simple relationship between a written text and its meaning. This is generally true in literature and particularly in poetry. A very simple experiment can be made in the classroom to demonstrate this point. Take a passage of about four hundred works from a book. Make a list of simple words like 'I', we, you, and, or, which, etc, which occur, showing the number of times each is found in the passage. Rewrite the passage omitting all these words and present the class with this together with the list of simple words; ask the class to rewrite the original passage. When they have agreed what it should be, compare it with the original. It is intereseting to try this experiment with both a prose passage and a poem.

WHAT IS THE VALUE OF INFORMATION?

To state what is the value of any given information in a quantitative sense, even in a specific context, is extremely difficult. In general terms we may define the value of information as the degree to which it can be used to remove the uncertainty from a situation. Thus consider the problem of delivering a letter with an incomplete address. What if you only know the city, say Paris? What if you know the street, Avenue Mozart, Paris, but do not know the house number? If any detail of the address is missing, there will be a degree of uncertainty which additional information may help to remove.

Certain specific information is often required to assist in the consideration of alternatives before decisions can be made. For example, a knowledge of the time will help in a decision as to when to leave home to catch a train or, in a classroom, to decide if there is a period of time sufficient to complete a planned project.

To assist in making a decision in this year's vacation or holiday it may be useful to have information about available package tours. Do they meet a family's requirements in terms of available time and money, opportunities for sunbathing, swimming, excursions and so on? Tour organisers study the market to find out the kinds of tours which people want. They then arrange

packages to meet these requirements as far as possible and in their advertised information to emphasise those aspects which they think will be appealing. The cautious buyer is aware that such information is often biased to produce a desired effect and makes allowances.

Information can also be biased in other situations. A child in a class may give an answer which he thinks is likely to please the teacher rather than say what he really thinks. In making business forecasts, managers tend to give over-estimates of requirements for man-power and money in the expectation of some reduction in their budgets. Research workers may look around for excuses to reject any information which seems to upset a promising theory which they are developing.

For the best decisions to be made it is important to have information based upon data which possesses sufficient accuracy and in a form that can be readily interpreted. Too much data may lead to a confused situation and increase the difficulty of making a decision; consequently it is essential to concentrate on the most significant facts.

We should also note at this point that it is largely when information is communicated that it becomes valuable. However, the recording and communication of information raises questions of representation and coding which are examined in the next chapter. Finally, the concepts already introduced may need deeper study to emphasise the variety of the fields of human activity in which they are important. In Appendix I these and related concepts are further discussed to help clarify the definitions and use of specialised terms.

III. REPRESENTATION AND CODING

Natural language is a medium for communications between human beings and, by suitable combination of generally understood words, a speaker can describe complex situations with many different shades of meaning. This mode of representing information through the spoken word developed to writing and later to printing. In course of time the progression from spoken to written word has produced many forms of representations which change with use and local custom.

Writing and printing are techniques which have enabled man to store and retrieve information external to the human memory. We have been accustomed to these means of transmitting and recording information since early childhood and do not normally question the rules on which this representation of information is based. At school, children learn how to read and write, to use letters and words and to enhance their knowledge of the rules of grammar. What children do, in fact, learn at school among other things are accepted ways of representing information, and of coding and interpreting data. The rules of coding, that is the *syntax*, are stated in phonetic dictionaries and text books of grammar; they have been developed and are accepted by the contemporary users of a continuously changing language.

The kind of representation used depends first of all on the practical needs of human beings. Natural language does not always provide the most efficient method for representing human experience. Thus information about a landscape or a human face is most effectively presented in photographs or paintings. Although photographs are extensively used to record information, paintings frequently fail to achieve this because few people have the artistic skill necessary to present data in this medium or the ability to interpret it. Moreover, there are many kinds of information unsuitable for representation in pictorial form. Paintings must not, however, be excluded here because today's knowledge of life in previous centuries is based on drawings and paintings which could not be replaced by the most extensive descriptions in natural language.

CHOICE OF CODE

When transmitting or processing data it is often necessary for the pictorial or verbal representation to be converted into code. The selection of a suitable

code depends on the way in which the code is expressed and on the purpose for which it has been defined. In the past, a particular choice of code may have been a deliberate consequence of what was known at a particular time. More often perhaps choices have been made in a random or arbitrary way or the original reasons for the choice can no longer be traced. For example, the symbols used for numbers are economical in the number of pen movements required for their formation, but knowledge of the development of the Hindu system of characters on which they are based is lost in antiquity. For the purpose of this booklet, finding the origin of a certain code is of little importance. More important is the fact that once a code is in use it is usually very difficult to change it at a later date, even if circumstances change.

Different communications media have different features which determine their suitability for a given code. Technical limitations often restrict the user to a single medium: for example, electromagnetic waves are the medium most suitable for communications between a space-craft and earth. Sometimes a variety of data representations occurs within a given medium – for example, music and numerical data are recorded in different ways on magnetic tape. Often the transmission of data requires a sequence of different codes as in a telephone call when a conversation is first recorded in human speech form and then converted to electrical signals. At the receiving end, the electrical signals are decoded to create sound waves. During transmission, the original information, whether represented as sound waves or electrical signals, cannot be extracted unless the rules governing the representation and coding are fully understood.

In general, the transmission and manipulation of data implies a whole chain of activity, each link requiring a particular code. Choosing different codes for the separate links in the chain implies as many recodings. The best choice of code for a particular link will achieve a balance between the efficiency of transmission and the extra effort required in recoding.

MINIMAL CODING

One of the main reasons for deciding upon a particular code is efficiency, and efficient coding is sometimes referred to as *minimal coding*. However, a minimal code cannot be defined until suitable criteria have been agreed. These could include the physical length of a written message, the time for transmission or the cost of data storage.

The Morse code is an example of minimal coding designed for electrical pulse transmission, each character being represented by a unique combination of dots and dashes. Characters are separated by spaces which are longer in duration than the spaces between the components of each character. There exists a great variety of possible codes based on the use of dots and dashes, but the criterion used by Morse to make his choice was the need to produce messages of minimum length using the English language, knowing the frequency with which letters occurred within English words. Codes were then selected so that the most frequent letters had the shortest codes, thus minimising the length of time taken to transmit a typical message.

Additional or *redundant** information is sometimes used in a code for specific purposes. For example, within a large data processing system, information is coded in as condensed a form as possible, using a minimal coding whenever appropriate, so that the overall processing cost is minimised. However, the chosen code may not be the theoretically possible minimum because of certain requirements, such as security, accuracy or ease of interpretation. In computer systems, extra symbols are frequently added to the code for internal checking purposes to protect against faults during processing.

COMPUTER CODES

When the decision has been made to manipulate data by some specific machine, a computer for example, then it is possible to specify the code in detail. In computer systems there are usually two levels of coding, one which enables man to communicate easily with the machine and the other, the internal code, which enables the machine to work most efficiently. For human understanding, data for computers is usually coded as a string of *symbols* drawn from the *data alphabet* of the machine. This data alphabet usually contains the alphabetic characters A to Z, the numerical digits 0 to 9, and a set of miscellaneous symbols such as punctuation marks, asterisks and so on. Keyboards incorporating these symbols, similar to those of typewriters, may be used to enter data for later processing by the computer. In a similar manner, output from the computer is arranged in a form which is easily understood by human beings, normally using the same symbols as for input.

*For a discussion of redundancy, see Appendix I.

The internal code of a computer depends on the designer of the electronic circuitry and has no direct interest for the user. The translation from the user's code to the internal code is performed automatically. The internal code is normally based on a two-state system, which is considered in more detail below.

EXAMPLES OF CODES

Let us describe in more detail one or two particular systems of coding.

AIRLINE BAGGAGE SYSTEMS

Airline checkpoints throughout the world have been allocated a three-letter identification code by the International Air Transport Authority to be used for the handling and routing of baggage. When a passenger checks in at a ticket desk, each item of baggage is labelled with a code representing the destination city or airport. Typical codes are AMS which represents Amsterdam in the Netherlands and LHR which represents London, Heathrow Airport. We can have 17,576 (i.e. 26 X 26 X 26) different permutations of three letters constructed from a data alphabet of 26 letters. This allows the representation of that number of different checkpoints. As there are less than 5,000 specified destination points within the domestic and international baggage system an effective code can be selected in many different ways.

Another approach might be to use just two letters and one digit which would still give an adequate selection of 6,760 (i.e. 26 X 26 X 10). This alternative coding is less helpful to the baggage clerks or handling agents because in order to check the routing of a bag the clerks would have to refer more often to a list to determine the meaning of each code. This would waste valuable time and also mean the possibility of human error through laziness or fatique, or through misreading the list. The advantage of using a totally alphabetic coding system is that a code can be selected for a particular destination which has the property of being a mnemonic. This helps the clerk to remember the codes more easily and reduces the need to use a reference list.

For machine processing, the three-letter destination code becomes a data item within the total information system. The item will be part of the passenger and flight records and will be used to pass information to destination airports

as well as to transit airports. The choice of code must therefore satisfy the limitations of the mechanical processor as well as the demands of the human user. In a computer-based flight information system the basic unit of data handled by the computer is usually a symbol drawn from a fixed data alphabet. A typical data alphabet contains 64 symbols and so a three-symbol code can be selected in 262,144 ways (i.e. 64 X 64 X 64). This gives rise to even greater scope for redundancy but the cost is small compared with the savings in human effort when using a mnemonic system.

TWO–STATE CODING

A simple method for coding information is as a pattern of two-state (or binary) symbols, for example 1 or =, / or X. In an electronic computer, a two-state representation of information is normally adopted* because of the simplicity of two-state electronic switching circuits, and the direct relationship between boolean algebra and the logic circuits which are fundamental to computer design. The symbols from the user's code are recorded automatically as patterns of binary symbols (bits) by a device similar to an electric typewriter. This pattern of bits may either pass directly into the computer or be stored externally on punched tape, punched cards or on magnetic media such as drums, discs and tapes. The computer user is not concerned with internal representation of his data in a two-state form but sometimes needs to check the coding of symbols on media such as paper tape or cards where the coding can be seen. Manufacturers of computer equipment, concerned with the overall processing efficiency of a particular computer system, use internal coding systems which the non-specialist may find difficult to understand. However, it is becoming less and less necessary to understand internal codes because computer technology is being orientated even more towards the requirements of the user and embraces the processing of information coded as pictures or as speech. Magnetic and optical character readers are able to recode data directly into binary form without human intervention at a keyboard. The selection of computer coding systems attempts to achieve a balance between the external human requirements and the internal cost effectiveness of electronic processing.

---

*This representation is sometimes described as a pattern of binary digits or *'bits'*.

IV THE ORGANISATION OF DATA

We have so far discussed how information can be represented and recorded in a coded form as data, thereby making it possible to process the data and perhaps to derive new data, which in turn may lead to new information. Before we can examine what data processing involves we must first consider some of the problems of organising or structuring data for a particular purpose.

Given certain basic data some may be easily derived but others may require much time and persistence. Thus to find a person's telephone number is relatively easy because the telephone directory has been arranged so as to facilitate this process. The requirement having been specified, a system of storing the data has been planned accordingly and the data has a particular structure. However, as we saw earlier, if you have a telephone number but want a person's address, such a directory is of little use because the data within has not been structured for that purpose. A different problem may arise if one has a car for sale and is seeking a purchaser. Many people may wish to buy it, but how are they to be contacted? Two possibilities exist; first to search through the 'car wanted' column in the newspapers; secondly, to advertise in the 'cars for sale' columns. But which papers will the seller see and which will be seen by the potential buyer? Here is a case where the data is poorly structured and so makes the task a difficult one.

DATA STRUCTURES

At this stage we need to consider what are the components of a data structure. Strings of symbols, which may include both alphabetic and numeric characters and other symbols indicating spaces, form what are known as *data items*. A set of items is said to constitute a *data record*, a group of which may, in turn, constitute a *data file*. This use of terms is analogous to the office situation where a person's name may be an item recorded on a medical record card and stored in a doctor's filing cabinet. Typically, a record contains items of data relating to a particular person, including his name, date of birth, weight and height. A file consists of a set of records which have a similar structure, and in this case, the set consists of records which contain data relating to each of the doctor's patients.

More complex collections of data, involving many files of varying structure and content, and accessible to a number of users, may constitute a *data base*.

ACCESS AND RETRIEVAL

It has already been stated that before a data structure can be planned it is necessary to examine how the data will be used. One common and important application is the retrieval of certain items, which may be selected according to a variety of criteria. We quote some examples:

1. From a file on the immunisation of children, it may be necessary to obtain a list of those who are due for a particular injection so that appointments with a doctor can be arranged and the necessary supplies of vaccine obtained.
2. A property agent may wish to provide a potential customer with a list of houses for sale, each with two bedrooms, with a garage, and within 10 kilometres of a town centre.
3. A society may keep a record of members, together with information on their subscription payments, with the aim of providing both an alphabetic list of members and an address list for subscription renewal notices.

When deciding on a data structure in such cases we must consider the ease with which the data can be brought up to date as a child receives an injection, a house is sold, or a new member joins the society. This is the process of up-dating an item, record or file.

In organising the data, attention must be given to its volume, the importance of speed of access, the uses for which the information may be required and the ease of up-dating. Often advantages in one respect must be weighed against disadvantages in the others. Suitable storage media have also to be selected from those available.

Various methods exist for organising records. In one method the records are arranged one after the other so that, to obtain a particular record it may be necessary to read, or at least scan, all the preceding records. This is a sequential arrangement. Another way is for each record to contain a 'key-word' or 'key-number' (or just 'key') which indicates the address where it is stored; it may then be obtained directly without any need to refer to other records. This makes it possible to have access to any record directly, not merely to the next one; such access is called random access. This distinction is illustrated by the difference between music on a tape and on a record. With tape one must pass through, albeit quickly, a section of tape before a

particular piece of music can be played. With a record, aided by the sequence of musical items listed on the cover, it is possible to move the head *directly* to the required track to obtain a particular recording.

RANDOM ACCESS

The storage of music on a magnetic medium is typical of the way information in general can be stored in coded form. In computers magnetic tapes are the most common external storage media when access to large amounts of data in sequential order is required. The computer device for random access storage also resembles a music record. Thus data can be stored on a revolving magnetic disc, or drum, and read by a moving head, which can very quickly be moved to a required track. Even more conveniently, one head can be allocated to each track.

Returning to the example of child immunisation, a list was required which would show which children needed immunisation, at what time and by which doctor; also the amount of vaccine needed by each doctor. How to prepare such a list and how to store it depends primarily on the total amount of data, i.e. the number of children that the health authorities will have to deal with. For very few children one may store all the data on a simple handwritten list. For a larger group one may use a file of typed cards with one card for each child, and a further stage may be the use of simple punched cards which can be sorted by machine. For a large number of children, say for all children in a big city, it may be better to put the data on punched cards which can then be read into a computer for further processing. However many children there are the form of the data is the same for each child and the record must contain at least the following items: name, address, types and dates of previous vaccinations, name of doctor. When searching through a file of such records in sequence one can easily imagine that a strict order of these items is important, indeed for any search done by machine a precise coding and ordering of the items in each record is essential.

KEY WORDS

Another problem may sometimes arise in connection with this file. Suppose a certain child comes for medical help and the doctor wants information from a single record. If the records are numbered and the record number is known,

there is no need for him to search through the whole file and he can have direct access to this one record. The important difference between records stored for sequential access and those stored for random access is that in the latter case each record must contain a 'key-number' or 'key-word', indicating the address at which the record is stored. For random access a certain record can only be found if the key-word which leads to the address of the record is given. The key-word must give unique reference to that address. This can be handled by an index or by an algorithm which creates the address as a function of the key-word. In the simplest case the key-word and the address are identical. In other cases some automatic way of obtaining the address from the key is required.

TREE STRUCTURES

Another method of organising data is illustrated by the problem of identifying a bird that has been observed. If several reference books are available the most appropriate one to use will be that which has a variety of lists using different criteria. In only one list does the full information about a bird appear. The other sets of records contain only key-words for the birds and these act as links to the full records.

This way of storing data so that the required item is reached by branching is called a tree structure. For example, if a bird has been seen close to the sea, then the list structured around habitat should be consulted followed by the lists on size, shape of foot, colonising, migration habits etc. In this way the observed data is used in stages until sufficient details have been considered to provide a unique identification.

Let us also follow up example (2) above. Selecting a house with a given number of bedrooms, with or without garage, within 10 kilometres of a given centre, provides a typical small problem for data processing. For a small community with only a few houses it can easily be handled either as a simple list or on a set of hand-written cards. For a larger community it may be advantageous to use a mechanical sorting procedure, analogous to the use of edge-punched cards with a needle to select cards for houses having the desired characteristics. However, in a community of a million people or more there will be hundreds of families looking for houses with varying requirements. For such large communities one master file is compiled containing data on all the houses available. Clearly picking out one house with several given features

from hundreds can no longer be done efficiently either by sequential or random access; a tree or link method may be used instead.

To illustrate this approach, suppose the community area is divided into sections and all the data on houses in one section is stored in a simple block of storage locations within the computer. This immediately avoids having to search the entire file sequentially for a house in a given section. Further, within the block of storage for a given section the houses can be classified by using other characteristics such as the number of bedrooms, although the best mode of such sub-divisions may depend on the type of community. In a different situation there might be greater advantage in organising the primary blocks of data on the basis of house costs. Another factor to be considered is the level of difficulty of up-dating the file i.e. making a space for data on a new house for sale, or deleting one which is sold. Again key-numbers will help to resolve this problem. In the category lists only the key numbers of the houses need occur, but the categories in which a particular number appears would also be included as part of the data associated with each keyed address. Deletion of information about a house which is sold and inclusion of information about a house newly available would then be simplified.

> To summarise, a data structure should be planned in order to use the information it contains effectively and with ease, not only to aid the user most effectively but also to minimise the cost of computer time involved in performing the search.

V DATA PROCESSING

In this final chapter we shall indicate some of the more important processes which may be carried out efficiently within a computer using data initially supplied to it in a suitable form. These processes will be described, not in terms of specific applications of the computer but rather to emphasise that the processing of data, whatever its original information content may have been, may be carried out in different ways, and for many different purposes. Not all these processes necessarily require a computer but in many cases the scale and complexity of the processing problem is such that it would be quite impractical to attempt it without one.

We select, for this brief survey, the following examples of data processing.

1. *THE TRANSFORMATION OF RAW DATA INTO STRUCTURED DATA*. Raw data may come from many sources and take many forms, such as

- experimental measurements
- data signalled from satellites
- meterological data
- data collected on forms or in questionnaires
- meter readings
- business transactions (mail-order, banking, retail purchases)
- printed text
- lists of names and addresses

All such data will be coded and have some kind of structure or format, however elementary, but as a preliminary to further processing it may be necessary to change its structure into one more suitable for storing in the computer and for its subsequent use. The structures used may be very varied:

tables, lists, arrays, trees, records and so on;

and the processes to produce these may be equally varied:

labelling, grouping, collating, sorting, merging and so on.

Sometimes the aim will be to eliminate all unnecessary detail and reduce the data to a convenient and manageable form, as in compiling statistics.

The incoming data may *already* be highly structured but not in such a form as to be easily used. A particular example is a computer program written in a language convenient to the user but which must be converted into a sequence of commands which the computer itself will accept and execute. This process is called program translation and compilation, and is based upon the application of clearly defined but often sophisticated algorithms.

2. *ANALYSIS OF DATA.* The processing of data frequently consists of straightforward analysis, either before or after a structuring process. Thus statistical analysis involving means, standard deviations, correlation coefficients or other statistics, is an example. If the amount of data is unwieldy a prior selection may also be necessary to provide samples for analysis.

A more complex example is 'picture processing', the analysis of data representing a picture such as that on a TV screen. The analysis may be directed towards identifying certain words or numbers, or other shapes contained within the picture.

3. *PROCESSING OF DATA IN RECORDS OF FILES.* When data has already been assembled in the computer store in the form of records or files there are many standard processes which need to be carried out and which account for a substantial part of the total use made of computers. These include:

- copying from one computer store to another, or printing out
- updating e.g. substitution and addition of data
- searching e.g. identification and retrieval of a particular record or file by branching search (as in the game of 'twenty questions')
- retrieval of selected data items from all records or a selection of records, followed by listing or analysis
- restructuring of files e.g. sorting into alphabetical order.

4. *ANALYSIS OF MODELS.* Many situations in real life can best be studied by setting up an appropriate model, involving a set of relations between entities. The data must then be adequate to define the structure of the model and its initial state, and the computer program must contain the algorithms which govern the processing to be carried out. Conclusions about the behaviour of the model, and hence hopefully of the real system, are reached by applying the set of relations to the given initial conditions as necessary.

Many models used for purposes of simulation contain probabilistic elements, and require hypothetical or random data to be supplied in order to study the functioning of the system. Examples are found in the flow of traffic, queueing problems, control of stocks in a warehouse, and in many games of chance. We might include also the drawing of patterns and other art forms. At a more complex level weather prediction is a field in which both deterministic and probabilistic models are important.

5. *DECISION–MAKING PROCESSING.* An essential aim of this type of data processing is to provide someone with information which will directly assist him in reaching a decision or taking some action. This might apply to a teacher, engineer, manager, financier, army commander or someone who is merely playing a game.

A preliminary to such decision-making may be a statistical or model analysis as already described. Some decisions may be clear-out, but in many situations a unique solution to the problem is not possible, and some criteria must be adopted to establish an optimum solution. This may be found only after a lot of processing using algorithms which incorporate these criteria. For example we have the problem of providing the appropriate number of telephone lines between two cities necessary to satisfy the expected demand, subject to a given acceptable delay.

Another form of optimisation process is illustrated by the assessment of a chess position, leading to a recommended move. This may involve a reassessment of the position after a large number of possible moves, or series of moves. As the criteria for evaluating a position can be complicated, methods of limiting the search may be essential, thus suggesting a *heuristic* rather than an algorithmic approach to the problem. This is an exploratory method in which a solution is arrived at by a process of guided trial and error, that is by assessing regularly the progress towards a desired result and calling a halt when the improvement is considered adequate. Heuristic processes may be valuable when algorithmic processes are too time-consuming.

This brief description of different kinds of data processing is not exhaustive but aims to show the variety of purposes, and different levels of complexity, for which processing may be carried out. It should be clear that for successful processing operations:

- data must be structured in accordance with the specific needs of the user;
- programs must be written which incorporate appropriate algorithms or other rules fos processing and organise the output from the system

We may repeat what was said at the start, that in principle there is little which is new or unusual in all these ways of processing data; what the computer makes possible is better-informed decision-making.

FINAL NOTE

Many important matters have been referred to but not explained in this booklet. The following unanswered question will be dealt with in further booklets:

- What is the computer system which carries out the processing, and what constraints does this place on what can be achieved?
- What are algorithms and algorithmic processes?
- How does one communicate with the computer system and write programs for it? How does a program control the processing?
- How is a computer used in specific applications, particularly those which may help the teacher?

APPENDIX I FURTHER EXPLANATION OF CONCEPTS

We shall assume that sources of information do exist. These sources are in nature or in human activity. Generally the meanings are complex; they can be considered as composed of *elementary meanings* using certain rules. The set of all complex meanings (or messages) is called the *semantic field*; and the set of all elementary meanings is called a *repertoire of semantic signs*.

To represent information we use a set of elementary symbols (letters in a wide sense), with which we can construct *complex symbols* (words, strings or expressions, in a wide sense) using certain rules. The set of elementary symbols is called an *alphabet*, and the rules used to make up expressions are called *syntactic rules*. The set of all expressions made from the alphabet using the syntactic rules is called *language*. A *grammar* is the system involving alphabets, and syntactic rules on them.

To provide communication of information, we need a relation between the semantic field and the language, with which we attach a meaning to each expression. Sometimes this relation is called a *code*, and we say that we code the semantic field. A complex meaning is then said to be *represented* in the language.

Sometimes coding is also understood as the replacement of one alphabet by another. Decoding is the inverse relation of coding.

Communication* is the transfer of information from one person to another, or data from one source to a receiver (human or mechanical). To make communication possible, information must be coded on a physical medium, thus we have to convert information to data, transmit this data through a channel and then decode the transmitted data.

Frequently during transmission, data is altered by random and/or unpredictable perturbations. This alteration is called *noise*. Hence, in decoding data as it arrives one may deduce information not conforming exactly to the originally coded information. As a general rule this does not lead to misinterpretation, but there is a possible danger of it doing so.

---

*Actually one communicates information, but transmits data; however, we shall use the word in the broader sense.

To prevent and correct errors produced by channel noise it is necessary to transmit the same string of data several times or to transmit some extra checking data even though redundant, i.e. the aim of that extra data is not to modify the meaning of the transmitted message but to protect it. Codes with redundant data for the detection or the correction of errors produced by noise during transmission are called *detecting* or *correcting* codes respectively.

The procedure for processing information generally is to act on data or expressions that carry information, knowing how the information and data are related. The *processor* is controlled by a special kind of data, carrying information in the form of an algorithm-encoded program. This processor, acting under the program's command, interprets the input data directly, processes it as required, and arranges for output of the data requested.

APPENDIX II SUGGESTED TOPICS FOR CLASS DISCUSSION

In any subject area there are topics which can illustrate interesting aspects of the concepts of information and information processing. We soon become aware that each separate discipline may contain its own specific information, structured in a particular way. Successful use of the computer in teaching that discipline will depend strongly on taking into account those characteristics of its information or data structure which are appropriate to it. Class discussion about these topics might seek answers to such questions as the following:

1. What information is involved?
2. What is its source?
3. How is it represented?
4. What processing of data is involved?
5. Who interprets the data?
6. What is the purpose of this information and its processing?
7. Can you design a set of questions to obtain information, or a form to record data in this topic, or a way of cataloguing it?
8. How can a computer be of use?

An example from cooking. It is usual to seek a recipe if you want to cook something for the first time. We may answer the above eight questions as follows:

1. The recipe will say what ingredients are required and how they should be put together and cooked to produce the dish. Several types of information are involved — how to get the best ingredients, what quantities for a given number of people, in what order to mix the ingredients, in what type of utensils they should be cooked and at what temperature.
2. The information comes from someone who has had successful experience in producing the dish — perhaps after much trial and error.

3. It may be written down in words and numbers or it may be dictated or demonstrated.

4. The recipe may need adapting to provide for a different number of people, to change from one set of units (pounds, pints) to another (metric), to adapt to different temperature scales or different working facilities.

5. In the first place the cook interprets the data but the 'proof of the pudding is in the eating'.

6. The recipe is written to enable cooks to provide this particular dish.

7. The set of questions would include quantities and probable costs of ingredients, a list of cooking facilities needed, food values, vitamin content, etc. Dishes could be catalogued alphabetically or by type (meat, fish, sweet, pastry).

8. A computer would seem to have little to offer in this situation but it could assist in providing another way of cataloguing, of updating prices, of calculating food values, or providing data information about dishes to satisfy various requirements of cost, taste or calorific content.

Further examples. The above questions may be considered in the context of the following subjects:

art (visual)

- painted portrait; a posed colour photographic portrait
- original painting; a first-rate copy
- abstract designs or patterns on materials
- designing a building, e.g. a house, hospital or factory

biology

- verbal description of a vegetable e.g. a carrot or cabbage
- categorisation by genus, variety, etc. of a plant, bird or animal
- arrangement of information about wild flowers or birds to help identification

- information about a special variety of a flower for a seed catalogue
- categorisation of soils
- categorisation of diseases of plants, animals
- disease diagnosis

cookery

- food values of ingredients
- designing a balanced diet
- different methods of working e.g. bake, steam, grill etc.
- best ordering of procedures in preparing a meal

economics

- value, price and cost
- power, labour, skill, know-how, relative values
- foreign exhange; variation of rates
- methods of taxation
- capital gains and losses
- inflation, cost of living, standard of living
- insurance; actuarial calculations

geography

- collection of information: physical, meteorological, geological, vegetation, industrial, population, ethnographic, social organisation, communications.
- representation in words, statistics and maps
- function of the surveyor
- town planning; distribution of facilities

- map making; codes in different types of maps showing contours, communications, crops, rainfall, barometric pressures.
- maps of a continent, ocean or other large region.

history

- genealogical tables
- 'trees' of events and consequences
- sources of information, manuscripts, authorities, tradition, archaeological discoveries; relative reliability.
- relations between different aspects of the history of a state e.g. economic, geographical and social.

household management/home economics

- information about size, shape and purpose of room; possible furniture and furnishings; fitting and matching needs
- information about methods for the routine running of a household
- organisation of shopping (stock control?)
- constraints (due to personal job routines etc.); flexibility
- design of entertainment and hospitality
- meeting needs for repair and maintenance; budgeting

language

- from spoken language to written language and vice versa
- from one spoken pronunciation to another (e.g. from another part of a country)
- from one language to another; spoken and written
- from one script (Latin) to another (Cyrillic, Chinese writing, Arabic)
- grammar, syntax and idioms in translation

- phonetic scripts and spoken language
- what is a word (sound, spelling and meaning)?
- different phases in learning to read: word and sentence recognition; word building; mechanical reading aloud; reading silently with comprehension

literature

- poetry as information — written and spoken
- drama as information — the parts played by author, producer, designer, director, actor, stage manager, audience
- information about settings — social, chronological, national
- requirements for a play or book to be approved by a particular critic; to be popular
- processing of information from author to reader or audience
- education in literature
- author identification: short word frequency, length of sentences etc.
- rules for constructing detective stories

mathematics

- information about a problem — in words, diagrams, tables of numbers, graphs, equations, inequalities
- patterns, sequences, iteration, algorithms
- intuitive and axiomatic geometry
- approximation, accuracy and errors
- random numbers and simulation
- statistical inference as data processing

music

- what does a musical composition 'mean' to a composer, conductor, performer, listener? — what data conveys the information?
- place of printed or written music
- sound mixing as information processing — is the information altered?
- mono and stereo
- music and dancing as coding
- how can computers and electronics be used in 'processing' music?
- what information is needed to specify the most popular records?

physical sciences

- information about the position of a body and about its shape, weight, the forces on it, its velocity, acceleration, kinetic and potential energy, temperature, colour
- changes of state, temperature
- different types of wave motion
- position and nature of stars and other astronomical objects
- nature of a substance, mixtures, compound — where found, how processed, how used; atomic structure
- natures of reactions — rates and types
- chain molecules; chemical reactions in living organisms

physical education

- identification of limbs, joints, muscles, arteries, organs
- effects of physical activity on various parts of the body — extension, contraction, flexing etc.
- design of activities to produce special effects — strengthening, giving greater mobility, etc.

- design of activities to give balanced effects on the whole body
- communication of intention from mind to muscle; data communication in the nervous system and brain
- development of physical skills, investigation and design of methods
- training for quicker reaction
- communication and processing of information in a ball game e.g. football or tennis; playing and refereeing
- rules for games; objectives, constraints
- football league tables; cricket or baseball personal statistics

social studies

- identification of different characteristics of groups into which people may be divided by racial origin, system of government e.g. wealth, ethical codes, living or working together, tastes and mobility
- identification of different priorities amongst different groups

miscellaneous

- methods of assessment; objectives of assessment; examinations, degrees, objective tests, interviews, projects etc.
- information required for time-tabling a school
- relevant information in choosing from a group of candidates for a job
- legal processes leading up to court action

ADDRESSES OF CONTRIBUTORS

Wm. F. Atchison (Chairman WG 3.1)
Dept. of Computer Science,
University of Maryland,
College Park, Maryland 20742 U.S.A.

A. Berger
Oesterreichisches Schulrechenzentrum
Spengergasse 20, A.– 1053 Wien
AUSTRIA

G.C. Bonham
Ministry of Education, Ontario
Nowat Block, Queen's Park
Toronto, Ontario
CANADA

U. Brondum
Rislundvej 3
8240 Risskov 2
DENMARK

R.A. Buckingham,
Birkbeck College,
Malet Street,
London WCIE 7HX,
U.K.

Sylvia Charp
School District of Philadelphia
Board of Education
5th and Lucerne Streets
Philadelphia, Pennsylvania 19140 U.S.A.

E. García Camarero
Centro de Cálculo de la Universidad
de Madrid
Av. Complutense S/N
Madrid – 3
SPAIN

J. Hebenstreit,
Ecole Supérieure d'Electricité,
Plateau du Moulon,
91190 Gif-sur-Yvette,
FRANCE

D.C. Johnson
Mathematics Education
330 Peik Hall
University of Minnesota
Minneapolis, Minnesota 5455 U.S.A.

R.E.J. Lewis
Chelsea Centre for Science Education,
University of London
Bridges Place, London SW6 4HR
U.K.

F. B. Lovis,
The Open University,
Walton Hall,
Milton Keynes, MK7 6AA
U.K.

J.F. Louw,
61 Revere
Zeederberg Street
Arcadia – Pretoria 0002
SOUTH AFRICA

T. Malmberg
Pedagogiska Institutionen
Ostra Agatan 9
S–753 22 Uppsala
SWEDEN

R.S. McLean
O.I.S.E.
252 Bloor St.W.
Toronto, Canada M5S 2H8

E.D. Tagg
Dept. of Mathematics
Cartmel College
University of Lancaster
Lancaster, LA1 4YL, U.K.

J.D. Tinsley H.M.I.
47 Goodby Road
Moseley
Birmingham B13 8RH,
U.K.

D.H. Wolbers
Delft University of Technology
Julianalaan 132, Delft
NETHERLANDS